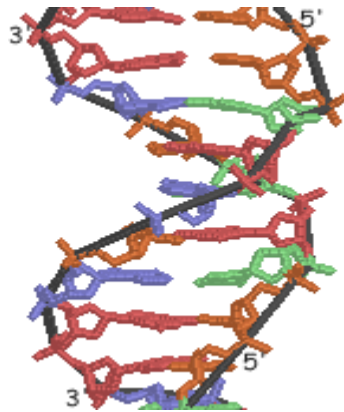


# **Er ustabile DNA-sekvenser samlokalisert med DNA-reparasjonsgener?**

- litteraturstudie og sekvensanalyse



Prosjektoppgave av Chris Lutken  
ved det medisinske fakultet,  
Universitet i Oslo,  
April, 2007

Veileder: Jarle Breivik  
leder ved forskerlinjen,  
Domus medica, Universitet i Oslo

<b><u>Innholdsfortegnelse:</u></b>	<b><u>side</u></b>
1.Abstract: .....	3
2.Innledning:	
2.1 Kreft .....	4
2.2 Det evolusjonere utgangspunktet for kreftutvikling.....	4
2.3 Fragile sites.....	4
2.4 Linkage disequilibrium og ekspansjon av ustabile sekvenser .....	5
2.5 Målet med oppgaven.....	6
3.Metoder: .....	6
4.Resultater:.....	8
5.Diskusjon:.....	10
6.Vedlegg (tabeller):.....	11
7.Referanseliste:.....	16

## 1. **Abstract:**

Fragile sites show susceptibility to DNA damage, leading to alterations that contribute to cancer development. They are associated with accumulation of unstable AT- and CCG-repeats which lead to unusual secondary structures and an increase in flexibility as well as instability. With the selfish gene hypothesis, which postulates that each gene tends to optimize its own replication, it seems like a paradox that these unstable sequences even exist because they represent an evolutionary disadvantage and should therefore be disfavoured by natural selection.

This disadvantage would be compensated for if these fragile sites were linked to or involved a repair gene that prevents instability/breakage. Fragile sites and repair genes should therefore be in linkage in the genome. It is basically two ways of understanding this: either that fragile sites infiltrate DNA repair genes, or that DNA repair genes have a mutation bias that causes expansion of unstable sequences in and around these genes. Either way one would think that these repair genes would have a high repeat content. The aim of this study was to make a map for linkage between DNA-repair genes and fragile sites as a preliminary study and then to scan these genes for CCG- repeat sequences.

## **2. Innledning:**

### *2.1. Kreft:*

Kreft er et resultat av akkumulering av mutasjoner som gir cellen enten nedsatt evne til apoptose (programmert celledød), manglende veksthemming, eller en kombinasjon av begge deler. Blant risikofaktorer som kan skape slike mutasjoner finner vi f.eks. røyking, diett og stråling samt at noen individer er disponert via arv. I en normal celledivisjon er det imidlertid mekanismer som hindrer disse forandringene i å skje. Med andre ord må det skje mutasjoner som hindrer cellens reparasjonsmekanismer og som gir cellen en vekstfordel i det gitte miljø. Eksempler på slike mutasjoner er delesjoner av tumor suppressorgener (som bremser uhemmet vekst av cellen), aktivering av proto-onkogener og amplifisering eller translokering av onkogener (som stimulerer cellen til å proliferere).

### *2.2 Det evolusjonistiske utgangspunkt for kreftutvikling:*

Biologisk evolusjon er genetiske forandringer som nedarves gjennom suksessive generasjoner. I en gitt populasjon vil det alltid finnes variasjon av genene til individene innen arten. Dette medfører at individene får forskjellige egenskaper som kan være enten en fordel eller en ulempe i ett gitt miljø. Individene med de fordelaktige egenskapene vil således utkonkurrere deres dårligere stilte "motstandere" og bringe sine egenskaper, og dermed gener, videre. Denne prosessen kalles naturlig seleksjon og er drivkraften i evolusjon. Naturlig seleksjon er gyldig for alle systemer av arv også for "programmet" for somatisk utvikling av organismen.

Gener hos encellede organismer selekteres etter deres evne til å få deres celle til å vokse så bra som mulig i et gitt miljø. Det er intet samarbeid med andre celler, og gener i en celle drar ikke noe fordel av gener i en annen celle. For multicellulære organismer er situasjonen en annen. Hos menneske skiller germinallinjen seg fra den somatiske allerede få dager etter at egg- og sædcelle har smeltet sammen. Her er gener i germinallinjen selektert for evne til å "bygge" og kontrollere en koloni av somatiske celler, som igjen skal sende germinale celler videre til neste generasjon. På denne måten har en altså fått selektert frem gener i germinallinjen etter deres evne til å holde kontroll på og stabilisere de somatiske cellene. Gener i den somatiske evolusjonslinjen kan derimot tenkes ha blitt selektert for deres evne til å få somatiske celler til å vokse fort og har ingen evolusjonistisk "grunn" til å bevare genetisk kontroll siden disse ikke skal føres videre til neste generasjon. Med dette som utgangspunkt

kan det virke uunngåelig at kreft med genetisk ustabilitet og ukontrollert vekst av celler utvikles etter hvert som vi eldres (figur 2).

### 2.3 *Fragile sites:*

Fragile sites er ikke-fargende gap/avstander på kromosomene som tidvis observeres når kromosomene er i metafasen. For å demonstrere dette er det vanligvis nødvendig å eksponere celler for kjemikalier som forandrer eller hemmer DNA-syntesen [1,2]. Man skiller vanligvis mellom sjeldne ("rare") fragile steder (FS) og vanlige ("common") FS. De sjeldne FS finnes i mindre enn 5 % av individer og induseres av folat mangel, mens de vanlige FS finnes hos individer flest og induseres av aphidicolin (inhibitor av DNA-polymerase) [13,7]. Flere forsøk har vist at FS gir DNA større fleksibilitet i disse områdene med risiko for forskjellige DNA-skader som f.eks. brudd på kromosomer med bl.a. delesjoner eller translokasjoner til følge [1,6,12,19]. Med fleksibilitet menes økt evne til å danne andre strukturer av DNA heliksen enn den vanlige beta-heliksen. Mekanismen bak fleksibiliteten ser ut til å være relatert til bestemte repeterte basesekvenser: CCG/CGG-"repeats" for de sjeldne FS og AT-"repeats" for de vanlige [3,14,7,9,16]. Man har funnet ut at CGG-repeats har evnen til å lage uvanlige sekundærstrukturer av DNA som f.eks "hairpin" [4], "slipped-strand" [15], eller quadruplex DNA [10] som hindrer eller forsinker DNA-replikasjon [17] og samtidig øker sannsynligheten for DNA-brudd. Dette gjelder også AT-repeterte sekvenser [8,11,21,22].

DNA-brudd kan igjen føre til translokasjoner, delesjoner eller amplifisering av reparasjonsgener, tumor supressorgener og proto-onkogener som er viktige "kontrollører" av normal cellyklus. FS er derfor en av flere faktorer som ligger til grunn for utvikling av kreft fordi disse stedene gjør kromosomene ustabile ved å føre med seg en tendens til brudd [20].

### 2.4: *Linkage disequilibrium og ekspansjon av ustabile sekvenser :*

Med en evolusjonistisk tankegang til grunn ville en kanskje trodd at disse fragile sekvensene som er ustabile ville blitt selektert bort gjennom årenes løp fordi de gir en ulempe (øket risiko for brudd og tap av genetisk materiale). Så hvorfor finnes disse sekvensene fortsatt i genomet vårt? En teori går ut på at de ustabile sekvensene nedarves koblet til stabile sekvenser gjennom linkage disequilibrium [26,27,29]. Sekvenser sies å være koblet (linkage) hvis de har en LOD-score på tre eller høyere hvilket i praksis tilsvarer en lengde på 10 CentiMorgan eller ca. 7,5 millioner basepar [30]. Slike sekvenser vil ha økt sannsynlighet for å nedarves sammen i forhold til to tilfeldige, uavhengige sekvenser. Hvis det er sann at de ustabile sekvensene nedarves sammen med stabile gener vil de ikke bli selektert bort fordi man da også måtte ha

selektert bort de essensielle genene.

Videre har man sett at noen gener har såkalte mutation bias [31,32]. Det vil si at disse genene vil ha bias mot for eksempel å gi insersjoner eller delesjoner i DNAet. Det tyder på at dette kan være tilfelle i en pågående studie av Mismatch repair (MMR) hvor der ser ut til at noen av disse reparasjonsgenene har en ekspansjons bias som øker lengden på de ustabile sekvensene, og at de selv har økt innhold av disse repeterte sekvensene [5].

Problemstillingen min var å se om dette gjaldt for flere reparasjonsgener. Særlig gjaldt det double strand break repair gener som er spesielt assosiert med "fragile site-skader"[18]. Både AT-repeats og CCG- repeats er assosiert med ustabilitet. I min oppgave har jeg valgt å fokusere på å lete etter reparasjonsgener med mutasjons bias for ekspansjon av CCG-repeterte sekvenser samt å se på hvor vidt disse er i linkage med fragile sites.

### 2.5 Målet med oppgaven:

For å finne ut av dette satte jeg opp hypotesene om at reparasjonsgener som har med fragile sites skader å gjøre burde ligge i eller i nærheten av disse stedene og da også ha en høy andel CCG-repeats i seg.

Målet med oppgaven var derfor: 1) å kartlegge hvilke steder (loci) på kromosomet reparasjonsgener ligger på og om disse ligger i ..  
nærheten av fragile steder.

2) Å analysere CCG/CGG-repeat innholdet i disse reparasjonsgenene samt å sammenligne disse med resten av genomet.

### 3. Metoder:

Problemstillingen gav meg en todelt oppgave: For det første måtte jeg finne informasjon om hvor de forskjellige fragile stedene og reparasjonsgenene lå. For det andre måtte jeg finne gensekvensene og telle antallet repeats i dem.

Jeg startet med å gå på pubmed (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi>) og søke på *search*: genome for "FRAGILE SITES" og fikk opp 2 svar på søket. Deretter skiftet jeg til *search*: GENE for fragile sites og fikk 133 treff på søket. Dette søket inneholdt imidlertid

andre arter enn det jeg var interessert i, og jeg forandret derfor til *search*: gene for "FRAGILE SITES HOMO SAPIENS". Denne gangen fikk jeg 126 treff og etter å ha bladd igjennom disse og skrevet ned de aktuelle, gikk jeg grundigere til verks og leste mer om hvert av de fragile stedene. Det var da i alt 120 stykker (noen av treffene var faktisk ikke fragile steder).

Deretter søkte jeg Pubmed for "REPAIR GENES HOMO SAPIENS" og fikk 460 treff. Etter å ha dobbeltsjekket disse genene ved å klikke på de på pubmed viste det seg at ikke alle var reparasjonsgener. I tillegg hadde et gen ofte flere navn slik at jeg, etter utsiling, satt igjen med 168 stykker. Ved å klikke på navnet på genet fikk jeg opp mer informasjon, for eksempel hvor langt genet er, hvor mange introner, eksoner genet har osv. Deriblant stod også locus til genet, og ved tilsvarende leting på de fragile stedene kunne disse kobles sammen med genene etter hvor de lå i forhold til hverandre. Resultatet ble satt på i en tabell i excel (tabell 3).

Neste trinn ble derfor å laste ned sekvensene til genomet samt de enkelte reparasjonsgenene. Til dette brukte jeg Ensembl (<http://www.ensembl.org/index.html>) hvor jeg søkte på og lastet ned FASTA-sekvensene til de relevante genene. Videre gikk jeg inn på downloads på ensembl, valgte Homo sapiens og lastet ned hele genomet. Det er bare en av trådene som blir lastet ned (gjelder både hele genomet og hvert enkelt av genene)- viktig med tanke på videre prosedyre og bruk av undernevnte program.

Deretter kjørte jeg de nedlastede filene gjennom et CCG-program [28] som telte lengden på de repeterte sekvensene (CCG/CGG). Programmet var laget slik at det talte totalt antall basepar og også antall "counts" dvs. antall baser i CCG-repeterte sekvenser. Eksempelvis ville en sekvens som ser slik ut: CCGCCGC gitt en "counts" på 7. En sekvens som denne: CCGCCGCCG ville gitt en counts på 6 (= CCGCCG) og en på tre (=den siste CCG) fordi den G'en på plass nr. 7 stopper tellingen (der skulle det vært en C for å passe inn i mønsteret). Legg merke til at her vil CGG tilsvare en CCG på komplementærtråden (bare en tråd ble lastet ned) og derfor kvalifisere indirekte som en CCG-repeat. Hvis en annen base som A eller T dukket opp i sekvensen stoppet også programmet opp og begynte å telle på nytt.

Resultatet av å kjøre sekvensene gjennom programmet var filer som kunne åpnes i excel som en tabell. Disse tabellene viste navn på genet, hvor mange baser genet har, hvor mange A,C,T og G baser, i tillegg til å vise hvor mange sekvenser det fantes med forskjellige antall counts. F.eks hvis genet så slik ut CCGCCGAAATTCGCGG ville CCG-programmet gitt en tabell med 6counts = 2stk. Programmet regnet også ut kumulative counts, dvs antallet baser summert opp for hver counts. Ett eksempel: hvis man hadde en 6counts (=6 baser) og en 8 counts ville det stått 6 under 6counts og 14 under 8 counts pga. summerte baser.

Neste stepp var å regne ut tetthet for de repeterte sekvenser i genene og genomet. Jeg tok da antall 2 counts delt på totale antall baser i genet, antall 3 counts delt på totale antall baser i genet osv. Dette resulterte i tetthetstall for hvert av genene og genomet. Jeg lagde en logaritmisk graf for å best kunne sammenligne resultatene (fig.1).

For statistisk analyse satte jeg opp nullhypotesen hvor jeg antok at CCG-repeats av lengde 12 og lenger er tilfeldig spredd utover i genomet. Jeg har da satt en cut off verdi lik 12 ettersom det bare er repeats på over 12 som er relevante med hensyn til fleksibilitet og sekundære strukturer ved CCG-repeterte sekvenser [25]. Hvis vi så antar at genomet har samlet lengde  $L_{\text{genom}}$  og mitt utvalg av gener har samlet lengde  $L_{\text{utvalg}}$ , vil sannsynligheten for at en tilfeldig valgt repeatposisjon ligger i en av de utvalgte genene være  $p = L_{\text{utvalg}}/L_{\text{genom}}$ . Antall CCG-repeats som havner i de utvalgte genene vil da følge en binomialfordeling  $\text{Binomial}(n, p)$  hvor  $n$  er totalt CCG-repeats(>12) i genomet. Formelen blir som følger:

$$P[X \geq x] = \sum_{x..n} C(n, x) * p^x * (1-p)^{(n-x)}$$

der  $x$  er forekomsten av CCG(>12) i mine reparasjonsgener

Denne metoden ser primært på antall CCG-repeats jeg fant; lengden tas kun hensyn til via terskelen for hvilke som telles med ( $\geq 12$ ).

For utregning av hvert enkelt av genene brukte jeg formelen over med Bonferroni-korreksjon av P-verdiene.

#### 4. **Resultater:**

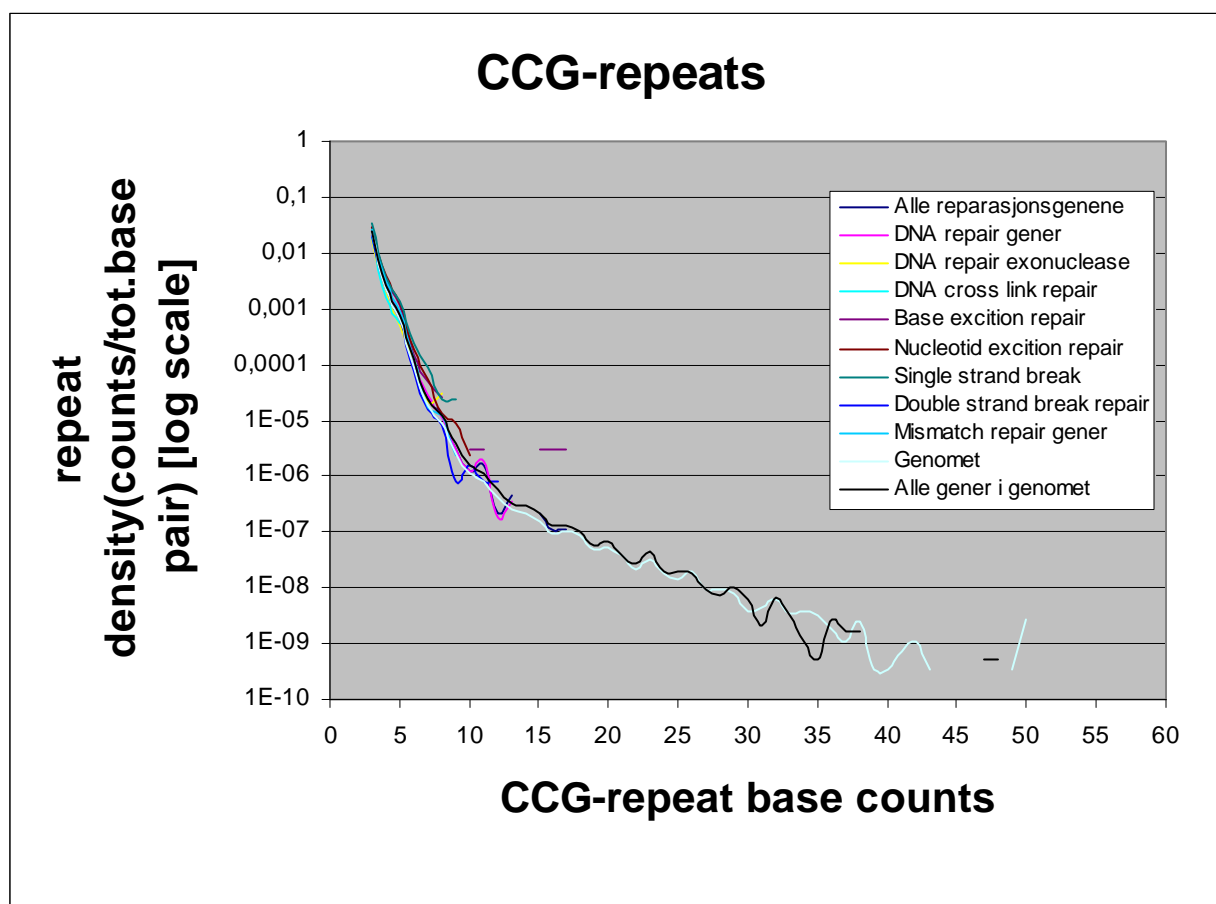
På kartleggingen av gener og fragile steder (se tabell 2) fant jeg at mange av reparasjonsgenene var på samme locus eller i nærheten av de fragile stedenes loci. Jeg lyktes ikke i å finne de bestemte sekvensene for de fragile stedene. Noen av reparasjonsgenene var ikke nærmere beskrevet slik at disse ble oppført som "DNA repair" (uten spesifisering av hva slags type DNA-skader de reparerer).

Når det gjelder **CCG**-repeat tetthet ser det, ut ifra grafen, ikke ut som innholdet av disse



sekvensene er forhøyet i forhold til resten av genomet. Noen av genene ser derimot ut til å ha et høyt kumulativt innhold med lange sekvenser (tabell 1). Av disse er det bare reparasjonsgenet ELN på kromosom 7 som ser ut til å være koblet til FRA7J sitt locus.

På den statistiske delen fant jeg ikke grunnlag for å støtte hypotesen om at CCG repeats forekommer hyppigere i reparasjonsgener. Sannsynligheten for å finne CCG-sekvensene i genene: PRKDC(double strand repair gene), HMGB1 (base excision repair), ELN, CSNK1D, CSNK1E, UBE2B, UBE2V1 og MAPK1 (alle DNA repair genes) var  $p = 0,59$ . Det eneste av disse genene med signifikant p-verdi var HMGB1 ( $p = 0,0011$ ).



**Fig.1:** Grafen viser tettheten av repeterte CCG-sekvenser i div. reparasjonsgener sammenlignet med genomet

## 5. Diskusjon:

Det var vanskelig å si noe generelt ut fra funnene, i og med at reparasjonsgenenes kurve ikke ligger konstant under eller over "genomkurven". Base-excision rep.genene er de eneste som ligger over genomkurven fra 15 counts og høyere. Det fantes ingen gener med counts på lenger enn 32 derfor er ikke counts på over 32 tatt med i tabellen.

Videre kan man se at noen av genene skilte seg ut så det kan også være interessant å se nærmere på disse enkelte genene. I forhold til CCG-repeats gjelder dette: PRKDC(double strand repair gene), HMGB1 (base excision repair), ELN, CSNK1D, CSNK1E, UBE2B, UBE2V1 og MAPK1 (alle DNA repair genes), men p-verdien = 0,59 ga ikke grunnlag for å forkaste nullhypotesen om at disse er tilfeldig fordelt.

HMGB1 sin p- verdi (0,0011) er såpass liten at det er sannsynlig at dette ikke skyldes en tilfeldighet. Dette genet koder for proteiner i gruppen "High Mobility Group" som er non-histone nukleære proteiner. Deres funksjon er todelt: En ekstracellulær cytokinivirkning og en intracellulær mer kompleks rolle. Sistnevnte består i å 1) binde seg til "minor groove" på DNA-tråden og dermed forvrengte denne til heliksstruktur, 2) regulering av transkripsjon og 3) reparasjon av feil DNA struktur som følge av f.eks. UV-stråling eller inkorporering av nukleotid analoger [33,34]. Det som er interessant i forhold til min studie er at HMGB1 ser ut til å ha direkte innvirkning på CCG repeats ved å fremme ekspansjon av disse [35].

Det er viktig å legge til at programmet som jeg brukte ble grundig testet underveis og i utviklingen av det, og jeg ser det som lite sannsynlig at det har telt feil. Hva angår resultatene av kartleggingen av rep.genene og de fragile stedene prøvde jeg å lete videre etter de bestemte basesekvensene - og ikke bare nøye meg med loci. Dette fant jeg på pubmed for rep.genene, men ikke for de fragile stedene så jeg fikk ikke sjekket denne koblingen noe utover det som er vist i tabell 2. Det er derfor vanskelig å si hvorvidt de fragile stedene er koblet til reparasjonsgenene eller ikke- selv når er på ca. samme locus.

Når det gjelder det statistiske, er det unøyaktig å sette opp en hypotese som antar at de repeterte sekvensene er tilfeldig spredt i genomet. For det første er ikke genet en tilfeldig sekvens, og for det andre opptrer ofte repeats i "blokker" slik at de ofte dukker opp flere i samme gen. Dette er spesielt viktig i forhold til HMGB1 hvor jeg fikk et signifikant avvik fra nullhypotesen. Det er neppe holdbart å anta at de fem repeatene er uavhengige forekomster.

Til slutt ser det ut til at videre forskning på CNG-repeterte sekvenser (hvor "N" er en hvilken som helst av basene C,G,A eller T) kan gi resultater [24].

## 6. Vedlegg/tabeller:

**Tabell 1:** Tabellen viser div. reparasjonsgener med høyt innhold av CCG-repeterte sekvenser. Genets navn er markert i grønt. Bp = basepar. Det vises lengden på den repeterte sekvensen samt antallet av denne type. Eks. for genet PRKDC er det én repetert sekvens inneholdende 12 baser.

### Double strand repair genes:

Repeat analysis for motifs GCC/CGG in file <b>PRKDC</b> .txt									
Gene	Bp(total)	repeatlengde	12	13	14	15	16	17	18
8 dna:chromosome chromosome:NCBI35:8:4884822 2:49035296:1	187075	Ant:	1	0	0	0	0	0	0

### Base excision repair

Repeat analysis for motifs GCC/CGG in file <b>HMGB1</b> .txt									
Gene	Bp(total)	repeatlengde	12	13	14	15	16	17	18
13 dna:chromosome chromosome:NCBI35:13:299308 84:30089731:1	158848	Ant:	0	2	0	1	1	1	0

### DNA repair

Repeat analysis for motifs GCC/CGG in file <b>ELN</b> .txt									
Gene	Bp(total)	repeatlengde	12	13	14	15	16	17	18
7 dna:chromosome chromosome:NCBI35:7:7288716 9:72928657:1	41489	Ant:	0	1	0	0	0	0	0

Repeat analysis for motifs GCC/CGG in file <b>CSNK1D</b> .txt									
Gene	Bp(total)	repeatlengde	12	13	14	15	16	17	18
17 dna:chromosome chromosome:NCBI35:17:777955 35:77824862:1	29328	Ant:	1	0	0	0	0	0	0

Repeat analysis for motifs GCC/CGG in file <b>CSNK1E</b> .txt									
Gene	Bp(total)	repeatlengde	12	13	14	26	27	28	29
22 dna:chromosome chromosome:NCBI35:22:370111 98:37119027:1	107830	Ant:	0	1	0	1	0	1	0

Repeat analysis for motifs GCC/CGG in file <b>UBE2B.txt</b>									
Gene	Bp(total)	repeatlengde	12	13	14	15	16	19	32
5 dna:chromosome chromosome:NCBI35:5:1337347 69:133755698:1	20930	Ant:	0	0	0	0	0	0	1

Repeat analysis for motifs GCC/CGG in file <b>UBE2V1.txt</b>									
Gene	Bp(total)	repeatlengde	12	13	14	15	16	17	18
20 dna:chromosome chromosome:NCBI35:20:481310 70:48203667:1	72598	Ant:	0	0	0	1	0	0	0

Repeat analysis for motifs GCC/CGG in file <b>MAPK1.txt</b>									
Gene	Bp(total)	repeatlengde	12	13	14	15	19	20	21
22 dna:chromosome chromosome:NCBI35:22:204414 27:20546284:1	104858	Ant:	0	0	0	0	0	1	0

**Tabell 2:** Tabellen viser kartleggingen av de fragile steder og reparasjonsgenes loci.

Gener og fragile steder som ser ut til å ligge på samme sted er farget i samme farge. DSB = double strand break repair, BER = base excision repair, NER = nucleotide excision repair, SSB = single strand break repair.

<u>Kromosom nr</u>	<u>Fragile sites</u>		<u>Repair genes</u>	<u>Locus</u>	<u>Function</u>	<u>Repair gene sequence</u>
1	FRA1A	1p36	TP73	1p36.3	Mismatch repair	3592286-3673013
	FRA1B	1p32	RAD54L	1p32	DNA repair	46425437-46456165
	FRA1C	1p31.2	USP1	1p31.3	DNA repair	62614700-62629430
	FRA1L	1p31	GADD45A	1p31.2-p31.1	DNA repair	67862904-67866040
			MSH4	1p31	Mismatch repair	75974651-76090944
			GADD45A	1p31.2-p31.1	DNA repair	67862904-67866040
			MSH4	1p31	Mismatch repair	75974651-76090944
	FRA1H	1q42.1	PARP1	1q41-q42	BER	222902526-222855241
			EXO1	1q42-q43	Mismatch repair	238337576-

						238379089	
2	FRA2A	2q11.2		REV1L	2q11.1-q11.2	DNA repair	99564998-99475932
	FRA2F	2q21.3		ERCC3	2q21	nucleotide-excision repair	127767982-127731096
	FRA2D	2p16.2		MSH6	2p16	Mismatch repair	47921937-47945743
3	FRA3A	3p24.2		RAD18	3p25-p24	DNA repair	8980146-8896561
	FRA3C	3q27		RFC4	3q27	DNA repair	188006992-187990386
4	FRA4C	4q31.1		HMGB2	4q31	BER	174630322-174627518
5	FRA5A	5p13		RAD1	5p13.2	DNA repair exonuclease	34954142-34944034
	FRA5C	5q31.1		RAD50	5q31	double strand break repair	131920529-132007498
6	FRA6B	6p25.1		WRNIP1	6p25.2	DNA repair	2710665-2730978
7	FRA7I	7q36		XRCC4	7q36.1	Double strand break repair	82409073-82685335
	FRA7B	7p22		PMS2	7p22.2	DNA mismatch repair	5821915-5786114
				NUDT1	7p22	DNA repair	2055098-2064021
				RPA3	7p22	DNA repair	7531478-7449815
	FRA7J	7q11		ELN	7q11.23	DNA repair	72887169-72927681
8	FRA8C	8q24.1		RAD21	8q24	DSB	117956182-117927355
	FRA8E	8q24.1					
10	FRA10A	10q23.3		POLL	10q23	NER	103337963-103328631
				DNTT	10q23-q24	DNA repair	98054202-98088311
	FRA10E	10q25.2		DCLRE1A	10q25.1	DNA cross-link repair	115603597-115584474
	FRA10F	10q26.1		MGMT	10q26	DNA repair	131224495-131455356
	FRA10B	10q25.2		MMS19L	10q24-q25	NER	99248155-99208073
11	FRA11A	11q13.3		HTATIP	11q13	DSB	65236065-

						65243651	
	FRA11H	11q13		RAD9A	11q13.1-q13.2	DNA repair	66915999-66922459
				UVRAG	11q13.5	DNA repair	75203923-75531342
				DDB1	11q12-q13	NER	60857125-60823502
	FRA11C	11p15.1		GTF2H1	11p15.1-p14	DNA repair	18300719-18345153
	FRA11I	11p15.1					

12	FRA12A	12q13.1		SMUG1	12q13.11-q13.3	DNA repair	52869024-52861519
	FRA12E	12q24		UNG	12q23-q24.1	BER	107998135-108011518
	FRA12C	12q24.2		POLE	12q24.3	DNA repair	131874295-131811626
	FRA12D	12q24.13		RAD9B	12q24.11	DNA repair	109402820-109432611
				GTF2H3	12q24.31	NER	122643261-122670214

14	FRA14B	14q23		MNAT1	14q23	DNA repair	60271301-60505137
	FRA14C	14q24.1		ALKBH	14q24	DNA repair	77244097-77209148
				ATXN3	14q24.3-32.2	NER	91642707-91599683
				RAD51L1	14q23-24.2	DNA repair	67356262-68132367
				MLH3	14q.24.3	Mismatch repair	74587886-74553053

16	FRA16A	16p13.11		ERCC4	16p13.3-p13.11	NER	13921524-13949705
				MPG	16p13.3	BER	69147-75841
				NTHL1	16p13.3	BER	2037868-2029817

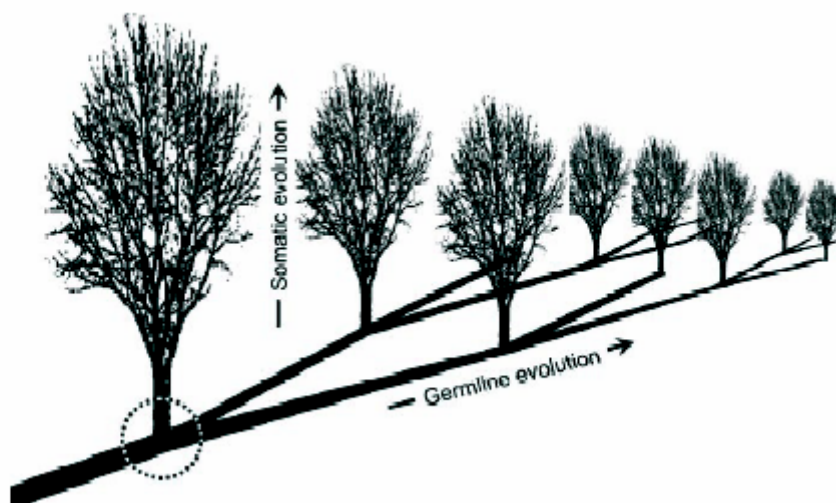
17	FRA17B	17q23.1		RAD51C	17q22-q23	DNA repair	54124962-54166691
				BRIP1	17q22-q24	DSB	57295537-57114767

19	FRA19A	19q13		XRCC1	19q13.2	SSB	48771555-48739304
				ERCC2	19q13.3	NER	50565669-50546686
				ERCC1	19q13.2-q13.3	NER	50619017-50604712
				PNKP	19q13.3-	NER	55062630-

				q13.4		55056274
			POLD1	19q13.3	DNA repair	55579408-55613082
			RUVBL2	19q13.3	DNA repair	54188968-54210994
			LIG1	19q13.2-q13.3	DNA repair	53365372-53310515
	FRA19B	19p13	RAD23A	19p13.2	NER	12917654-12925455
			XAB2	19p13.2	NER	7600439-7590412

20	FRA20A	20p11.23		XRN2	20p11.2-p11.1	DNA repair,exonuclease	21231952-21318463
----	--------	----------	--	------	---------------	------------------------	-------------------

X	FRAXE	Xq28		TREX2	Xq28	DNA repair	152231025-152256892
	FRAXF	Xq28					



**Figur 2:** Det evolusjonære utgangspunktet for kreft. Multicellulærhet innebærer en repetitiv divergens mellom germinallinje- og somatisk evolusjon. Gener i germinallinjen er selektert for deres egenskaper til å holde kontroll og stabilitet i de somatiske cellene. Når de går inn i de somatiske rekkene derimot, er genene selektert etter evne til å vokse raskt. Utfall av vekstkontroll og genetisk ustabilitet viser seg da som en uunngåelig konsekvens og

opprinnelsen til kreftutvikling i det evolusjonistiske skillet mellom germinal- og somatisk linje.

## 7. Referanseliste:

- [1] Thompson and Thompson: Genetics in medicine. 6<sup>th</sup> edition 2004
- [2] Arlt MF, Casper AM, Glover TW.: Common fragile sites. Cytogenet Genome Res. 2003;100(1-4):92-100.
- [3] Finnis M, Dayan S, Hobson L, Chenevix-Trench G, Friend K, Ried K, Venter D, Woollatt E, Baker E, Richards RI.: Common chromosomal fragile site FRA16D mutation in cancer cells. Hum Mol Genet. 2005 May 15;14(10):1341-9. Epub 2005 Apr 6.
- [4] Gacy AM, Goellner G, Juranic N, Macura S, McMurray CT: Trinucleotide repeats that expand in human disease form hairpin structures in vitro. Cell. 1995 May 19;81(4):533-40.
- [5] Breivik J., Falster D.: simple sequence repeats accumulate within the DNA mismatch repair genes that protect them from deletion. Molecular biology and evolution. 2005 24.nov
- [6] Mimori K, Druck T, Inoue H, Alder H, Berk L, Mori M, Huebner K, Croce CM.: Cancer-specific chromosome alterations in the constitutive fragile region FRA3B. Proc Natl Acad Sci U S A. 1999 Jun 22;96(13):7456-61.
- [7] Sarafidou T, Kahl C, Martinez-Garay I, Mangelsdorf M, Gesk S, Baker E, et al.: Folate-sensitive fragile site FRA10A is due to an expansion of a CGG repeat in a novel gene, FRA10AC1, encoding a nuclear protein. Genomics. 2004 Jul;84(1):69-81.
- [8] Fukuda H, Katahira M, Tanaka E, Enokizono Y, Tsuchiya N, Higuchi K, Nagao M, Nakagama H.: Unfolding of higher DNA structures formed by the d(CGG) triplet repeat by UP1 protein. Genes Cells. 2005 Oct;10(10):953-62.
- [9] Sutherland GR, Baker E, Richards RI.: Fragile sites still breaking. Trends Genet. 1998 Dec;14(12):501-6.
- [10] Fry M, Loeb LA.: The fragile X syndrome d(CGG)<sub>n</sub> nucleotide repeats form a stable



- tetrahelical structure. *Proc Natl Acad Sci U S A*. 1994 May 24;91(11):4950-4.
- [11] Jackson JA, Trevino AV, Herzig MC, Herman TS, Woynarowski JM: Matrix attachment region (MAR) properties and abnormal expansion of AT island minisatellites in FRA16B fragile sites in leukemic CEM cells. *Nucleic acid Res*. 2003 Nov 1; 31(21):6354-64.
  - [12] Mishmar D, Mandel-Gutfreund Y, Margalit H, Rahat A, Kerem B.: Common fragile sites: G-band characteristics within an R-band. *Am J Hum Genet*. 1999 Mar;64(3):908-10.
  - [13] Glover TW, Berger C, Coyle J, Echo B.: DNA polymerase alpha inhibition by aphidicolin induces gaps and breaks at common fragile sites in human chromosomes. *Hum Genet*. 1984;67(2):136-42.
  - [14] Matsuyama A, Shiraishi T, Trapasso F, Kuroki T, Alder H, Mori M, Huebner K, Croce CM.: Fragile site orthologs FHIT/FRA3B and Fhit/Fra14A2: evolutionarily conserved but highly recombinogenic. *Proc Natl Acad Sci U S A*. 2003 Dec 9;100(25):14988-93. Epub 2003 Nov 20.
  - [15] Pearson CE, Wang YH, Griffith JD, Sinden RR.: Structural analysis of slipped-strand DNA (S-DNA) formed in (CTG)<sub>n</sub>. (CAG)<sub>n</sub> repeats from the myotonic dystrophy locus. *Nucleic Acids Res*. 1998 Feb 1;26(3):816-23.
  - [16] Yu S, Mangelsdorf M, Hewett D, Hobson L, Baker E, Eyre HJ, Lapsys N, Le Paslier D, Doggett NA, Sutherland GR, Richards RI.: Human chromosomal fragile site FRA16B is an amplified AT-rich minisatellite repeat. *Cell*. 1997 Feb 7;88(3):367-74.
  - [17] Samadashwily GM, Raca G, Mirkin SM.: Trinucleotide repeats affect DNA replication in vivo. *Nat Genet*. 1997 Nov;17(3):298-304.
  - [18] Schwartz M, Zlotorynski E, Goldberg M, Ozeri E, Rahat A, le Sage C, Chen BP, Chen DJ, Agami R, Kerem B.: Homologous recombination and nonhomologous end-joining repair pathways regulate fragile site stability. *Genes Dev*. 2005 Nov 15;19(22):2715-26.
  - [19] Finnis M, Dayan S, Hobson L, Chenevix-Trench G, Friend K, Ried K, Venter D, Woollatt E, Baker E, Richards RI.: Common chromosomal fragile site FRA16D mutation in cancer cells. *Hum Mol Genet*. 2005 May 15;14(10):1341-9. Epub 2005 Apr 6.
  - [20] Popescu NC.: Genetic alterations in cancer as a result of breakage at fragile sites. *Cancer letters* 2003; 192:1-17
  - [21] Zlotorynski E, Rahat A, Skaug J, Ben-Porat N, Ozeri E, Hershberg R, Levi A, Scherer SW, Margalit H, Kerem B.: Molecular basis for expression of common and rare fragile sites. *Mol Cell Biol*. 2003 Oct;23(20): 7143-51

- [22] Handt O, Sutherland GR, Richards RI.: Fragile sites and minisatellite repeat instability.: Mol Genet Metab. 2000 Jun;70(2):99-105.
- [23] R.Dawkins, The Selfish Gene (oxford university press, new york. ed.2nd ,1976)
- [24] Paiva AM, Sheardy RD.: The influence of sequence context and length on the kinetics of DNA duplex formation from complementary hairpins possessing (CNG) repeats. J Am Chem Soc. 2005 Apr 20;127(15):5581-5
- [25] Paiva AM, Sheardy RD.: Influence of sequence context and length on the structure and stability of triplet repeat DNA oligomers. Biochemistry. 2004 Nov 9; 43(44): 14218-27
- [26] M.Ackermann, Lin Chao.: DNA sequences shaped by selection for stability. PLoS Genet. 2006 Feb;2(2):e22.
- [27] Goldstein D.: Islands of linkage disequilibrium. Nature genetics 2001;29:109-111
- [28] Program av Daniel S. Falster; Department of Biological Sciences, Macquarie University, Australia
- [29] K.Ardlie,L. Kruglyak, M.Seielstad : Patterns of linkage disequilibrium in the human genome. Nature Genetics 2002 Jun 21;29(5576):2225-9
- [30] Thompson and Thompson: Genetics in medicine 6<sup>th</sup> edition 2004; 8: 118-128
- [31] M. Gomes-Pereira,D. Monckton.: Chemical modifiers of unstable expand simple sequence repeats: What comes up, could come down. Mutat Res. 2006 Jun 25;598(1-2):15-34.
- [32] C. Savouret, E. Brisson: CTG repeat instability and size variation timing in DNA repair-deficient mice. EMBO J. 2003 May 1;22(9):2264-73
- [33] Takeshi Suda , Yukio Mishima et al: A novel activity of HMG domains: promotion of the triple-stranded complex formation between DNA containing (GGA/TCC)<sub>11</sub> and d(GGA)<sub>11</sub> oligonucleotides. First Department of Biochemistry, Niigata University School of Medicine, Japan,1996.
- [34] S. Müller, L. Ronfani, M. E. Bianchi: Regulated expression and subcellular localization of HMGB1, a chromatin protein with a cytokine function. Journal of Internal Medicine volume 255 Issue 3 Page 332 - March 2004.
- [35] Zhao Y, Cheng W, Gibb CL, Gupta G, Kallenbach NR. : HMG box proteins interact with multiple tandemly repeated (GCC)<sub>n</sub> (GGC)<sub>m</sub> DNA sequences. J Biomol Struct Dyn. 1996 Oct;14(2):235-8.